



Published in final edited form as:

*Qual Life Res.* 2016 March ; 25(3): 615–624. doi:10.1007/s11136-015-1131-3.

## Evaluating PROMIS® instruments and methods for patient-centered outcomes research: Patient and provider voices in a substance use treatment setting

Kelly L. Johnston<sup>1</sup>, Suzanne M. Lawrence<sup>1</sup>, Nathan E. Dodds<sup>1</sup>, Lan Yu<sup>1,2</sup>, Dennis C. Daley<sup>1</sup>, and Paul A. Pilkonis<sup>1</sup>

<sup>1</sup>Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213, USA

<sup>2</sup>Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA

### Abstract

**Purpose**—Our work as a primary research site of the Patient-Reported Outcomes Measurement Information System (PROMIS®), combined with support from the Patient-Centered Outcomes Research Institute, allowed us to evaluate the real-world applicability and acceptability of PROMIS measures in an addiction medicine setting.

**Methods**—As part of a 3-month prospective observational study, 225 outpatients at a substance abuse treatment clinic completed PROMIS item banks for alcohol use (as well as 15 additional item banks from 8 other PROMIS domains, including emotional distress, sleep, and pain), with assessments at intake, 1-month follow-up, and 3-month follow-up. A subsample of therapists and their patients completed health domain importance ratings and qualitative interviews to elicit feedback regarding the content and format of the patients' assessment results.

**Results**—The importance ratings revealed that depression, anxiety, and lack of emotional support were rated highest of the non-alcohol-related domains among both patients and clinicians. General alcohol use was considered most important by both patients and clinicians. Based on their suggestions, changes were made to item response feedback to facilitate comprehension and communication.

**Conclusions**—Both therapists and patients agreed that their review of the graphical display of scores, as well as individual item responses, helped them to identify areas of greatest concern and was useful for treatment planning. The results of our pilot work demonstrated the value and practicality of incorporating a comprehensive health assessment within a substance abuse treatment setting.

---

Correspondence to: Kelly L. Johnston.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical standard** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was reviewed and approved by the University of Pittsburgh Institutional Review Board.

## Keywords

Patient-reported outcomes; Substance abuse; Qualitative research; Interviews

---

## Introduction

The role of patient-reported outcomes (PROs) in comparative effectiveness research (CER) has expanded in recent years as recognition of their usefulness has grown [1, 2]. Incorporating PROs into research can help determine not only the efficacy, but the effectiveness of a given treatment, particularly in situations in which there is not a clear winner based on clinical variables or survival [3, 4]. In a recent meta-analysis evaluating the clinical impact of PROs, Marshall, Haywood, and Fitzpatrick concluded that including such measures in clinical settings improves the diagnosis and management of patient conditions, enhances patient-provider communication, and may positively influence patients' health status. These findings were particularly evident in the realm of mental health [5].

A common methodological challenge for CER is identifying assessments of health status and health-related quality of life relevant to key stakeholders (e.g., patients and health care providers). Optimally, such instruments should be brief; be easy to read, administer, score, and interpret; and be meaningful in the clinical encounter [6]. The increased awareness of the need for research with which to guide clinical decisions, as well as the benefits of PROs, has led to several federally funded initiatives. One such program is the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>). Established in 2004, PROMIS is an NIH Roadmap initiative devoted to developing better tools for assessing constructs relevant to the investigation and treatment for all chronic diseases (e.g., pain, emotional distress, sleep) [7–21]. Using a domain-specific approach rather than a disease-specific approach, PROMIS involves patients and content experts to inform the content within each domain. By utilizing modern psychometric methods (item response theory; IRT) and computerized adaptive testing (CAT), the PROMIS item banks are able to assess these constructs with greater precision and less patient burden than traditional methods.

Another initiative, the Patient-Centered Outcomes Research Institute (PCORI), was established in 2010 under the auspices of the Patient Protection and Affordable Care Act, with the mission of identifying research priorities, establishing an agenda, and carrying out CER [22]. A primary aim of PCORI was to promote research that is informed by the perspectives, interests, and values of patients and other health care stakeholders throughout the research process [23]. In the current study, we combined our work as a PROMIS research site with the PCORI agenda. Using a combination of quantitative and qualitative techniques, we tested longitudinally our PROMIS item banks for alcohol use (as well as 15 other item banks from 8 other domains of health) in an outpatient addiction medicine clinic serving patients with “dual diagnoses,” i.e., both substance use and psychiatric disorders.

In keeping with established PROMIS methodology, the item banks for alcohol use were generated through qualitative item analysis (including focus groups with both substance use patients and social drinkers from the community), reviewed by content experts, and evaluated quantitatively using techniques from both classical test theory (CTT) and IRT, as

described by Pilkonis et al. [24]. The PROMIS methodology has been described in greater detail previously [8, 10, 24, 25]. The alcohol item banks include five subdomains: alcohol use, negative consequences, negative expectancies, positive consequences, and positive expectancies. See Table 1 for sample items from each subdomain.

The primary goal for this report was to evaluate the acceptability and ease of use of the PROMIS measures when administered within a substance use treatment setting. Assessments were completed within a month of the beginning of a substance use treatment episode (T1), and approximately 1 and 3 months later (T2 and T3) to ultimately examine responsiveness to change within a typical treatment episode. To achieve the main goal, we conducted interviews with a subsample of patients and their clinicians at T1 and T3 about their experience of completing the PROMIS measures and the value of receiving feedback on the health domains that they assessed. We included the five subdomains of the alcohol banks and 15 other item banks from 8 other PROMIS domains to form a comprehensive PROMIS health status profile. A goal of this profile was to identify potential predictors of outcome, including comorbid conditions such as depression, anxiety, and sleep disturbance which are common among individuals with substance use disorders [27]. However, comprehensive assessment of physical and mental health may not always be included systematically in their care. The efficiency of the PROMIS measures, especially via CAT, makes it feasible to generate a rich health status profile at a small cost of time and effort (usually just 4–6 items for each domain).

In summary, our work used the health profile to align the PROMIS and PCORI agendas and to integrate PRO assessments within the context of clinical practice. This integration has the potential to influence the process of care, particularly in the detection and management of mental health conditions as well as improvement in patient-provider communication [5, 28, 29]. However, assessment research embedded in the clinical treatment session has encountered many challenges, including difficulties in interpreting scores from PROs, intrusion on already brief clinical encounters, and the need for support at the technical, patient, and provider level to maintain integration [3, 30]. Within the larger aim of demonstrating the acceptability of the PROMIS measures in a clinical setting, our goals for the interview component were to learn which health topics were the most important to substance users and their clinicians, to determine the best means to convey that information to patients and clinicians, and to identify barriers to the implementation and sustainability of incorporating these procedures into an addiction medicine setting.

## Method

### Sample

Men and women 18 years and older who were able to read and understand English were enrolled in the main protocol. Participants were required to have recently begun outpatient treatment ( < 30 days) for substance use at the Center for Psychiatric and Chemical Dependency Services (CPCDS). A part of Addiction Medicine Services in the Department of Psychiatry at the University of Pittsburgh Medical Center, CPCDS specializes in the treatment for patients with both substance use and mental health problems, i.e., “dual-diagnosis” patients. Only participants who indicated that they had had a drink of alcohol

within the past 30 days were enrolled in the study because three of the alcohol item banks (alcohol use, negative consequences, and positive consequences) have a 30-day time frame. In addition, patients had to have attended at least one treatment session after their initial assessment to be eligible for the second and third assessments. Informed consent was obtained from all participants.

Table 2 provides a complete summary of the demographic characteristics of the interview sample. Our interview sample consisted primarily of men (65.8 %), with 39.7 % African American and 6.8 % Hispanic. The mean age of the sample was 39.1 (SD = 10.7), and 27.3 % had a high school diploma or less. A majority self-reported diagnoses of depression (84.9 %) and/or anxiety (80.8 %) in their lifetime. The interviewed patients had a mean AUDIT score of 29.2 (SD = 5.1). Our interview subsample was representative of the larger computer sample, which was comprised of slightly less men (56.0 %) and had slightly lower educational attainment (39.1 % had a high school diploma or less).

### Test administration

Assessments were completed within a month of the beginning of a substance use treatment episode (T1), and approximately 1 and 3 months later (T2 and T3). Of the initial 225 patients who completed T1, 165 completed T2 and 159 completed T3. A treatment dropout range of 26–60 % can be anticipated within the first month of treatment in a substance use population, and our attrition rate of 29.3 % was near the lower bound of this range [31–34]. In addition to treatment dropout, we also experienced attrition due to challenges inherent in a substance use population (e.g., participants were incarcerated or had relapsed).

Participants completed assessments on computers reserved for this purpose at CPCDS. Demographic information and self-reported medical, tobacco, and drug use histories were collected at T1. Participants were administered the CAT versions of PROMIS measures, including physical functioning, pain, fatigue, sleep/wake disturbance, emotional distress (depression, anxiety, and anger), cognitive functioning, sexual functioning, and alcohol use at all time points (see Table 1 for sample items). Legacy measures were also administered at each time point to demonstrate convergent validity with the PROMIS alcohol item banks. These included the Alcohol Use Disorders Identification Test (AUDIT) [26], a measure commonly used to indicate drinking of some clinical significance. Items were displayed one at a time using Assessment Center<sup>SM</sup>, the PROMIS electronic testing platform (see <http://www.assessmentcenter.net>). Content was randomized by block (e.g., alcohol use, other health domains), and measures were randomized within each content block. Randomization was used to reduce the risk of (a) participants becoming fatigued and responding in a less valid way to items at the end of the assessment and (b) participants having the experience of responding in a repetitive way to the same items from assessment to assessment.

### Generation of reports

Generating patient reports entailed a three-step process: (1) downloading participant data from the Assessment Center platform, (2) restructuring the data for use, and (3) assembling the data for presentation. The first step, downloading the data, was done using a web-based user interface provided within the Administration tab in Assessment Center's researcher

portal. Two data files, Assessment Data and Assessment Scores, were required to generate the reports. The contents of the Assessment Data file were parsed by a Statistical Analysis System (SAS) software [36] process to remove unnecessary variables. The variables retained included the participant's identifier, assessment time point, item identifier, and the corresponding response score. These data were used to generate item-level reports displaying the respondent's scores for each of the items answered. The Assessment Scores data file was then processed by Microsoft Access [37] using structured query language to create a graphical display of the participant's T-scores for each domain, presented in a fixed order originally proposed by the research team (but later changed to reflect the preferences reported by interview respondents). The procedures used, including scripts, query language, and a guide to running the protocol, are available upon request.

Providing participants with responses to individual items was only feasible because we had used CAT versions of all of the item banks. Because the average item bank only required five items ( $SD = 2$ ), we were able to include all assessed items in a report. This kept the report relatively short and readable.

## Interviews

In keeping with PCORI's aim of capturing patients' experiences in their own words, qualitative interviews were conducted at T1 and T3 to elicit feedback from patients on two versions of the cross-sectional and longitudinal results from the computerized assessment. See Fig. 1 for examples of the two versions of the graph format. All interviews were conducted by master's level clinicians and were recorded using digital voice recorders. In general, interviews for both patients and clinicians took 30–60 min to complete. See Table 3 for the total number of interviews completed as related to the number of unique clinicians. Since some clinicians were treating more than one of our interview patients, they were invited to participate in an interview for each unique patient, representing a range of 1–7 patients per clinician. Informed consent for the interviews was provided by participants and their clinicians separately from the larger study protocol.

## Patient interviews

Participants whose total scores on the AUDIT were 20 or greater were contacted via phone or in clinic by study staff and asked to participate in an interview for which they would receive additional compensation. Scores of 20 or more on the AUDIT warrant further evaluation for alcohol dependence [35]. By requiring a minimum score of 20 for interview participants, we were able to focus on patients with clear alcohol issues. Interviews were not conducted at T2 to decrease participant and clinician burden.

Interviewees were shown the reports detailing the results of their PROMIS CAT assessments in both bar graph form and as a table of individual items and responses, and asked to give feedback regarding the validity and acceptability of the information. Interview questions were open-ended, with further probes and clarifications as needed. The interviews focused on whether interviewees thought the information would be helpful in treatment (e.g., would it make their visits more efficient, improve patient-provider communication), their comfort level in discussing the information with their therapist, the relevance of the content domains

to their treatment, and what they liked or disliked about the formats of their assessment results. Additionally, and perhaps most importantly, their comments and suggestions for the improvement of the formats were solicited. These patients were again invited to participate at T3.

### **Clinician interviews**

The clinicians for all interviewees were also invited to participate in concurrent interviews, which included a review of their patients' responses and scores. Clinicians were invited to participate in an interview for each unique interview patient they were treating. Clinician interview probes were similar to those in the patient interviews, but from a treatment provider's perspective. Questions elicited information about the general usefulness of the reports in a therapy setting as well as their relevance for the specific patient in question. Clinicians were also asked for their feedback regarding format and their suggestions for improvement.

### **Ratings of importance**

Following the interview, patients and clinicians were asked to rate the importance of the health domains assessed. Each domain was rated on a scale of 1 (*less important*) to 10 (*more important*). The same ratings were also completed for the alcohol use item banks separately. Patients provided individual ratings. However, as described above, clinicians provided scores for multiple patients at baseline. Therefore, we computed an average domain score for each clinician to create a profile of importance for clinicians. To compare the mean ratings of importance, we performed a one-way (within-subjects) analysis of variance with post hoc comparisons among the means.

### **Interview review process**

Patient and clinician interview responses were entered into a database, including open-ended responses and the numerical ratings. The research team met approximately once a month to discuss interview responses and suggestions for improvements offered by patients and clinicians. If a particular suggestion to change appeared more than once, it was reviewed by the research team and considered for implementation. Changes were included in new versions of the output if they were deemed feasible within the software capabilities, suggested by more than one participant (patients or clinicians), and endorsed by the research team. Additionally, interviewer observations were incorporated into the revision decision-making process outside of direct participant suggestions (e.g., comprehension issues). Patients' and clinicians' responses regarding the utility and relevance of the assessment results and the domains covered were coded as yes/no responses and tabulated for the sample.

## **Results**

In this study, participants answered 4–6 items per CAT and a median of 94 CAT items total. The 20 administered PROMIS CATs had a median completion time of 17.0 min.

Patients and clinicians found the information to be helpful and relevant to treatment, and patients indicated that they would be comfortable discussing the information with their therapists. Responses to questions pertaining to helpfulness, relevance, and comfort ranged from 91 to 100 % positive for both versions 1 and 2. Both the bar graph and the table of items and responses were changed in response to feedback. See Table 4 for a detailed description of the changes to the graph.

### Ratings of importance

Among patients, the mean ratings of importance reflected three tiers that emerged from post hoc comparisons (see Table 5). The domains in the first tier of importance were those for emotional distress and social isolation (designated by letters A and B). The largest, second tier included a mix of sleep difficulties, impairments in social participation, cognitive issues, and some physical limitations (letters C, D, and E). The third tier included issues related to pain and sexuality (letter F). For the alcohol use item banks, patients rated alcohol use, negative consequences, and negative expectancies as most important (letters A and B), with positive consequences and expectancies of lesser importance (letter C).

Among clinicians, overall ratings of importance tended to be higher, in general, than for patients. The range of means for clinicians on the 10-point scale was 9.1–5.0 (with a median of 7.7) versus a range of 8.8–4.1 (with a median of 6.4) for patients. The rank order of the domains, however, was quite similar, with one exception: clinicians regarded anger ( $M = 8.1$ , rank = 5) as a more important issue than patients ( $M = 6.1$ , rank = 10). For the alcohol use item banks, clinicians also showed more concern about the perceived positive aspects of alcohol use (both consequences and expectancies) than patients. With only 13 clinicians rating variable numbers of patients, however, most of the post hoc comparisons of their mean ratings of importance were not statistically significant, given the small  $n$  and large standard errors. Nonetheless, the results from clinicians alert us to potential differences between their perceptions and those of patients.

### Comprehension issues

Some domain names were changed to improve comprehension (e.g., “cognitive concerns” became “concentration problems,” and “social roles” became “managing daily responsibilities”). Domains in which a high score indicated better functioning were reverse scored and relabeled to be in concordance with the other domains so that high scores indicated greater severity of symptoms (e.g., “managing daily responsibilities” was changed to “unable to manage daily responsibilities”, and “emotional support” was changed to “lack of emotional support”). In other words, high scores were aligned to indicate poorer functioning across all domains.

### Graphical display

Several participants and clinicians suggested adding color to the graphs to more clearly indicate that higher scores signaled a problem to investigate further. In response, a color version was created with a shading of green on the low, less severe end and orange on the high, potential problem end. To increase comprehension of the scores displayed, we included indicators at T-scores (the standard metric of PROMIS scales) of 40 and 60, a brief

sentence on the graph itself about the indicators, and the actual numerical score next to the relevant bar on the graph. Scores between 40 and 60 (the range from  $-1$  SD to  $+1$  SD on the T-score metric) are considered normative scores in the general population. We also created a Frequently Asked Questions document that was made available to both patients and clinicians (available upon request).

Other significant formatting changes included displaying all domains on a single page. We also standardized the order of the domain names as opposed to listing them from the highest to lowest scores. This change ensured that patients and clinicians would see related item banks in relevant groupings (e.g., all domains for emotional distress—depression, anxiety, and anger—were clustered together). No significant changes were suggested by participants or clinicians, or observed by the research team, for version 2. See Fig. 1 for a comparison between sample versions 1 and 2.

### Item report

To preserve space on the original individual item format, only the participant-chosen response was displayed. However, after feedback, a header of all possible responses was displayed. Numerical anchors (i.e., response scores) were removed due to the reverse scoring indicated above. The time frame was also included to orient participants to the temporal context of their responses.

We also assessed participants' preference for the graphical format or item display and received mixed feedback. Whereas some patients and clinicians preferred the easy-to-read graphical display, others preferred to view the individual items. We chose to preserve both formats to accommodate individual preferences and learning styles.

### Discussion

One of the main goals of our work was to incorporate stakeholders throughout the process. As described above, our primary stakeholders were patients with substance use disorders. Their clinicians, along with our content experts, were included as secondary stakeholders. Our project goals were to identify the most salient health constructs for these patients and to create the best formats to integrate information about these constructs into a substance use treatment setting.

Because we solicited ratings of importance across multiple domains, we were able to identify the most critical health topics to be included in a substance use treatment setting. As might be expected in a dual-diagnosis setting (with substance use and psychiatric diagnoses), depression and anxiety were rated highest by both patients and clinicians. Lack of emotional support and social isolation was also ranked in the top tier. Social support has been shown to be a strong predictor of success in treatment and therefore warrants attention in treatment sessions [36–40]. The second tier of importance included a blend of several domains (sleep difficulties, impairments in social participation, cognition, and physical limitations), all of which were rated as moderately important factors. The final tier, related to issues of pain and a lack of interest in sex, was rated as less important, likely due to the younger mean age of the sample.



The ordering of the importance of the domains was similar in both clinicians and patients (see Table 5), with the exception of anger, which clinicians rated as a more important issue. This may be due to a combination of patients distancing themselves from a stigmatized emotion and clinicians identifying it more readily than patients as a part of ongoing therapeutic sessions. Also, the PROMIS alcohol item banks reflecting positive domains (positive consequences and expectancies) were rated lower by patients than clinicians, presumably because individuals in treatment for substance use have identified their drinking as a problem and are encouraged to think of it negatively. Substance use treatment providers may place more emphasis on the positive domains out of concern that their patients are still attaching positive beliefs and consequences to their alcohol use, which may make them more vulnerable to relapse.

A second goal was to determine the best formats in which to present health profile information in a substance use clinical setting. Based on patient and clinician suggestions, we made several changes to our original graph and item-by-item formats (as represented in Table 4; Fig. 1). Our updated versions were well received by both patients and clinicians, and we did not identify any further problems to address. We also appreciated the need to present the information in two different formats (text and graphical) to accommodate individual preferences and learning styles. By incorporating patient and provider feedback, we were able to provide a more patient-centered experience and generate user-friendly, visually appealing, consistent, and easy-to-understand displays endorsed directly by patients and clinicians.

Regardless of the appeal of feedback from the PROMIS health profile, an important question remains regarding its impact on treatment outcome: What evidence do we have that it enhances the effects of treatment, e.g., with lower rates of attrition or larger effect sizes? Preliminary analyses of treatment outcome in the current sample (using the PROMIS alcohol use CAT as the dependent variable) document that patients receiving the qualitative interview, with feedback on their health status profiles, did achieve better outcomes over the course of treatment (Pilkonis and colleagues, in preparation). Our current work, however, did not use a randomized, controlled design, and a more robust demonstration of the impact of feedback would be valuable to justify routine clinical use of such feedback.

Assessment research is notoriously difficult to integrate into the clinical encounter [29, 30], and several possible barriers to implementation were identified through the current study. In the current study, research staff needed to be available at the clinic to adequately accommodate the schedules of patients and clinicians and to generate the interview reports. Ideally, clinic staff or clinicians themselves would administer the assessments, but given staffing needs and time concerns this may not always be possible. In this study, we utilized desktop computers in private rooms to this purpose, but utilizing tablets may alleviate this concern. Alternatively, incorporating the assessment into patient web portals would give patients the flexibility to choose the time and location. Also, our health profile is generated from a CAT assessment that yields an IRT-calibrated score, and many electronic medical record systems do not have the capabilities for CAT algorithms. A possible compromise is the use of brief, static short forms that generate raw scores which can be transformed into IRT-calibrated scores.

While we appreciate the need for additional validation of the impact of feedback, we recommend the use of the PROMIS assessments within a clinical treatment setting. Both patients and clinicians valued the information generated from the PROMIS measures captured in a health profile and indicated that they would feel comfortable utilizing it in treatment sessions. Combining PROMIS methods with the PCORI agenda allowed us to make our project truly patient-centered. We must give patients a voice in the development process to provide valuable health information that can be understood, accepted, and integrated into treatment.

## Acknowledgments

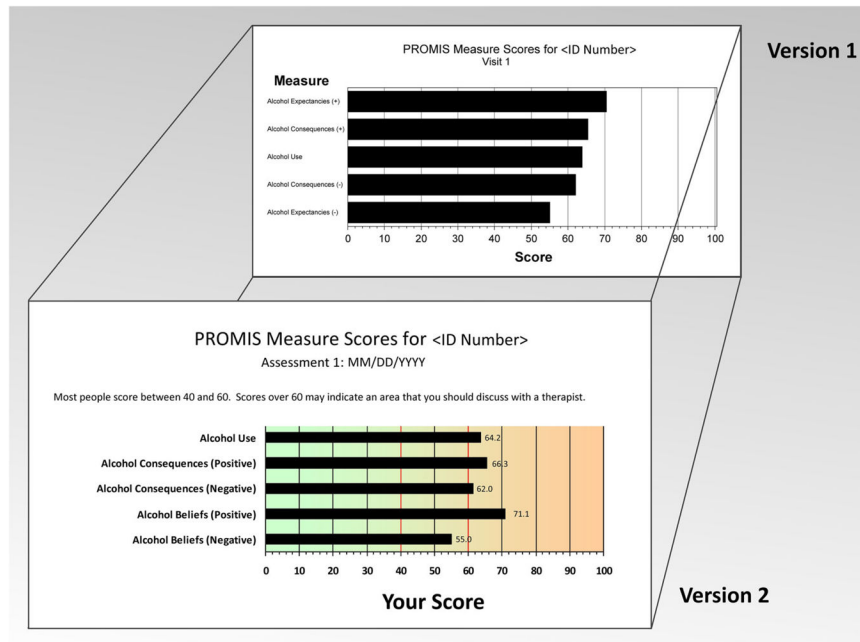
This project was funded by the Patient-Centered Outcomes Research Institute (1P2PI000189; PI: Paul A. Pilkonis, PhD). We acknowledge the work of our colleagues within the NIDA Clinical Trials Network that collaborated in this work. Dennis Daley, PhD, is the principal investigator of the Tri-State Appalachian node, and clinical data from this node were collected in Pittsburgh, PA, where we are grateful for the help of Trey Ghee, LSW, Dorothy Sandstrom, MS, and Janis McDonald.

## References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders 5. Arlington, VA: American Psychiatric Association; 2013. Assessment Measures; 733-748
2. U.S. Government Accountability Office (GAO). Comparative effectiveness research: HHS needs to strengthen dissemination and data-capacity-building efforts: GAO-15-280. 2015. [www.gao.gov](http://www.gao.gov)
3. Snyder CF, Jensen RE, Geller G, Carducci MA, Wu AW. Relevant content for a patient-reported outcome questionnaire for use in oncology clinical practice: Putting doctors and patients on the same page. *Quality of Life Research*. 2010; 19:1045–1055. [PubMed: 20424920]
4. Wu AW, Kharrazi H, Boulwar LE, Snyder CF. Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *Journal of Clinical Epidemiology*. 2013; 66(8 Suppl):S12–S20. [PubMed: 23849145]
5. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: A structured review. *Journal of Evaluation in Clinical Practice*. 2006; 12(5):559–568. [PubMed: 16987118]
6. Rotgers F. Clinically useful, research validated assessment of persons with alcohol problems. *Behaviour Research and Therapy*. 2002; 40(12):1425–1441. [PubMed: 12457637]
7. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: Item banking, tailored short forms, and computerized adaptive assessment. *Quality of Life Research*. 2007; 16(Suppl 1): 133–144. [PubMed: 17401637]
8. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system: Depression, anxiety, anger. *Assessment*. 2011; 18(3):263–283. [PubMed: 21697139]
9. Broderick JE, DeWitt EM, Rothrock N, Crane PK, Forrest CB. Advances in patient reported outcomes: The NIH PROMIS measures. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2013; 1(1):1–7. Article 12.
10. Buysse D, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010; 33(6):781–792. [PubMed: 20550019]
11. Revicki D, Chen W, Harnam N, Cook K, Amtmann D, Callahan LF, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*. 2009; 146(1–2):158–169. [PubMed: 19683873]
12. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *Journal of Rheumatology*. 2009; 36(9):2061–2066. [PubMed: 19738214]

13. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Maihoefer C, Lawrence SM. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>) in a three-month observational study. *Journal of Psychiatric Research*. 2014; 56:112–119. [PubMed: 24931848]
14. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS). *Medical Care*. 2007; 45:S22–S31. [PubMed: 17443115]
15. DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, et al. Construction of the eight item PROMIS pediatric physical function scales: Built using item response theory. *Journal of Clinical Epidemiology*. 2011; 64(7):794–804. [PubMed: 21292444]
16. Varni JW, Stucky BD, Thissen D, DeWitt EM, Irwin D, Lai JS, et al. PROMIS Pediatric Pain Interference Scale: An item response theory analysis of the Pediatric Pain Item Bank. *Journal of Pain*. 2010; 11(11):1109–1119. [PubMed: 20627819]
17. Amtmann D, Cook K, Jensen MP, Chen WH, Choi SW, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. *Pain*. 2010; 150(1):173–182. [PubMed: 20554116]
18. Flynn KE, Jeffrey DD, Keefe FJ, Porter LS, Shelby RA, Fawzy MR, et al. Sexual functioning along the cancer continuum: focus group results from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Psycho-Oncology*. 2011; 20(4):378–386. [PubMed: 20878833]
19. Forrest CB, Bevans KB, Tucker C, Riley AW, Ravens-Sieberer U, Gardner W, et al. Commentary: the Patient-Reported Outcome Measurement Information System (PROMIS) for Children and Youth: Application to pediatric psychology. *Journal of Pediatric Psychology*. 2012; 37(6):614–621. [PubMed: 22362923]
20. Hahn EA, DeWalt DA, Bode RK, Garcia SF, DeVellis RF, Correia H, et al. New English and Spanish social health measures will facilitate evaluating health determinants. *Health Psychology*. 2014; 33(5):490–499. [PubMed: 24447188]
21. Hays RD, Bjorner JB, Revicki RA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items. *Quality of Life Research*. 2009; 18(7):873–880. [PubMed: 19543809]
22. Patel K. Health reform's tortuous route to the Patient-Centered Outcomes Research Institute. *Health Affairs*. 2010; 29(10):1777–1782. [PubMed: 20921475]
23. Patient-Centered Outcomes Research Institute. 2011–2015 from [www.pcori.org](http://www.pcori.org)
24. Pilkonis PA, Yu L, Colditz J, Dodds NE, Johnston KL, Maihoefer C, et al. Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>): use, consequences, and expectancies. *Drug and Alcohol Dependence*. 2013; 130:167–177. [PubMed: 23206377]
25. DeWalt DA, Rothrock N, Yount S, Stone AA, on behalf of the PROMIS Cooperative Group. Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*. 2007; 45:S12–S21. [PubMed: 17443114]
26. Saunders JB, Aasland OG, Babor TF, Delafuente JR, Grant M. Development of the Alcohol-Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption. *Addiction*. 1993; 88:791–804. [PubMed: 8329970]
27. Stinson FS, Grant BF, Dawson DA, Ruan WJ, Huang B, Saha T. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug and Alcohol Dependence*. 2005; 80(1):105–116. [PubMed: 16157233]
28. Jensen RE, Rothrock NE, DeWitt EM, Spiegel B, Tucker CA, Crane H, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Medical Care*. 2015; 53(2):153–159. [PubMed: 25588135]
29. Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: Challenges and opportunities. *Quality of Life Research*. 2009; 18:99–107. [PubMed: 19034690]

30. Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: An overview and practical examples. *Quality of Life Research*. 2009; 18:125–136. [PubMed: 19152119]
31. Graff FS, Morgan TJ, Epstein EE, McCrady BS, Cook SM, Jensen NK, Kelly S. Engagement and retention in outpatient alcoholism treatment for women. *American Journal of Addictions*. 2009; 18(4):277–288.
32. Loveland D, Driscoll H. Examining attrition rates at one specialty addiction treatment provider in the United States: a case study using a retrospective chart review. *Substance Abuse Treatment, Prevention, and Policy*. 2014; 9(1):41.
33. Substance Abuse and Mental Health Services Administration Treatment Episode Data Set. Discharges from substance abuse treatment services Rockville, MD: Substance Abuse and Mental Health Services Administration; 2010 BHSIS Series S-67, HHS Publication No. (SMA) 14-48172013
34. Stark MJ. Dropping out of substance abuse treatment: A clinically oriented review. *Clinical Psychology Review*. 1992; 12(1):93–116.
35. Babor, TF., Higgins-Biddle, JC., Saunders, JB., Monteiro, MG. The alcohol use disorders identification test: Guidelines for use in primary care World Health Organization; 2001 (No. WHO/MSD/MSB/01.6a)
36. Tracy EM. Social support: A mixed blessing for women in substance abuse treatment. *Journal of Social Work Practice in the Addictions*. 2010; 10(3):257–282. [PubMed: 20953326]
37. Chou KL, Liang K, Sareen J. The association between social isolation and DSM-IV mood, anxiety, and substance use disorders: Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry*. 2011; 72(11):1468–1476. [PubMed: 21295001]
38. Warren JI, Stein JA, Grella CE. Role of social support and self-efficacy in treatment outcomes among clients with co-occurring disorders. *Drug and Alcohol Dependence*. 2007; 89(2–3):267–274. [PubMed: 17329040]
39. Richter SS, Brown SA, Mott MA. The impact of social support and self-esteem on adolescent substance abuse treatment outcome. *Journal of Substance Abuse*. 1991; 3(4):371–385. [PubMed: 1668229]
40. Ciraulo DA, Piechniczek-Buczek J, Iscan EN. Outcome predictors in substance use disorders. *Psychiatric Clinics of North America*. 2003; 26(2):381–409. [PubMed: 12778840]



**Fig. 1.** PROMIS profile graph versions 1 and 2

**Table 1**

## Sample items from PROMIS alcohol use item banks

---

Alcohol use
I spent too much time drinking
I drank too much
Negative consequences
I worried when I drank
I felt angry when I drank
Negative expectancies
People have trouble thinking when they drink
People feel sick the day after drinking
Positive consequences
My future seemed better when I drank
I was able to express myself better when I drank
Positive expectancies
People are outgoing when they drink
People have more fun at social occasions when they drink

---

**Table 2**

## Demographics and clinical characteristics

Characteristic	Interview sample ( <i>n</i> ) ( <i>n</i> = 73)	%
Sex		
Male	48	65.8
Ethnicity		
Hispanic	5	6.8
Race		
American Indian/Alaska Native	1	1.4
Asian	0	0.0
Black/African American	29	39.7
Native Hawaiian/Pacific Islander	0	0.0
White	41	56.2
Other/Multiracial	2	2.7
Education		
High school diploma or less	20	27.3
Further educational attainment	53	71.6
Mean age (SD)	39.1 (10.7)	
Income per year		
<\$20,000	52	71.2
Between \$20,000 and \$49,999	17	23.3
Between \$50,000 and \$99,999	2	2.7
\$100,000 or more	3	4.1
AUDIT score (SD)	29.2 (5.1)	
Polysubstance use (past 30 days)	53	72.6
Self-reported lifetime diagnoses		
Depression	62	84.9
Anxiety	59	80.8

**Table 3**

## Interview sample sizes

	Version 1		Version 2	
	T1	T3	T1	T3
Patient interviews	58	28	16	7
Clinician interviews	33	19	11	7
Unique clinicians <sup>a</sup>	11	9	7	6

<sup>a</sup>Clinicians were interviewed more than once



**Table 4**

## Graph format changes

<b>Original graphical format</b>	<b>New graphical format</b>
Black and white/gray scale	Additional color version (green to orange)
No normative score indicators	Dotted lines between scores of 40 and 60 to indicate normative scores
No description on graph	Brief sentence to describe scores outside of 40 and 60
No FAQs	Addition of FAQ document
Domains listed in the order of highest score by individual participant (may be in different order at different sessions)	Domains listed in specific, consistent order at each assessment time point
Original scoring used (and separated by types of scores)	Domains with high scores indicating good functioning were reversed scored, and all domains appeared on one page
Original domain names used:	Names changed due to rescoring and/or feedback:
Social roles	Unable to manage daily responsibilities
Social activity	Lack of social activity
Emotional support	Lack of emotional support
Cognitive concerns	Concentration problems
Interest in sex	Lack of interest in sex
Physical function	Limited physical function

Table 5

Domain ratings of importance by respondent

Respondent	Domain	Mean importance rating	Post hoc comparisons
Patient	Depression	8.8	A
	Anxiety	8.1	A B
	Lack of emotional support	7.4	A B C
	Social isolation	7.1	B C D
	Limited social activity	6.9	C D
	Sleep problems	6.7	C D
	Sleep-related impairment	6.6	C D
	Unable to manage daily responsibilities	6.4	C D E
	Concentration problems	6.1	C D E
	Anger	6.1	C D E
	Fatigue	5.9	C D E
	Limited physical function	5.8	D E
	Pain behavior	4.9	E F
	Pain interference	4.8	E F
	Loss of interest in sex	4.1	F
	Alcohol use	9.5	A
	Alcohol consequences (negative)	8.6	A
	Alcohol expectancies (negative)	7.5	B
	Alcohol expectancies (positive)	6.3	C
	Alcohol consequences (positive)	6.0	C
Clinician	Anxiety	9.1	
	Depression	9.0	
	Lack of emotional support	8.5	
	Social isolation	8.1	
	Anger	8.1	
	Limited social activity	7.9	
	Sleep-related impairment	7.7	
	Unable to manage daily responsibilities	7.7	

Respondent	Domain	Mean importance rating	Post hoc comparisons
	Sleep problems	7.6	
	Fatigue	7.1	
	Concentration problems	6.3	
	Pain behavior	5.9	
	Pain interference	5.9	
	Limited physical function	5.6	
	Loss of interest in sex	5.0	
	Alcohol use	9.2	
	Alcohol expectancies (positive)	9.1	
	Alcohol consequences (positive)	9.1	
	Alcohol consequences (negative)	8.4	
	Alcohol expectancies (negative)	8.2	

For the patient ratings of importance, means not sharing a common letter are significantly different ( $p < .05$ ) in post hoc comparisons using the Sidák method for multiple comparisons. With multiple clinicians rating a variable number of patients, the standard errors for clinicians' ratings are large, and most of the multiple comparisons are not statistically significant despite the mean differences